Transitioning from HDF4 to HDF5

Robert E. McGrath (mcgrath@ncsa.uiuc.edu)

Kent Yang (ymuqun@ncsa.uiuc.edu)

NCSA

University of Illinois, Urbana-Champaign

Important Note

- **Both HDF4 and HDF5 are supported by the NCSA HDF group.**
- **We will continue to maintain HDF4, as long as we are funded to do so.**
- **We recommend using HDF5, and that you consider migrating from HDF4 to HDF5 to take advantage of the improved features and performance of HDF5.**

See: http://hdf.ncsa.uiuc.edu/h4-h5.html

Main Points

The transition from HDF4 to HDF5 will require effort.

NCSA is working to assure that the transition is as smooth as possible

Two main points in this talk:

- Key technical challenges – 3 basic necessities
- Recent NCSA activities – building tools to meet these needs

I. Key Technical Challenges

The transition to HDF5 involves several technical challenges:

- Support both HDF4 and HDF5 (and HDF-EOS w/4 and HDF-EOS w/5)
- Interoperate HDF4 and HDF5 files and libraries
- Convert data from HDF4 to HDF5

We are working to provide tools and advice to make this as easy as possible.

1. Supporting both HDF4 and HDF5

- Will need to support two file formats, libraries, documentation, etc. for many years.
  - NCSA is committed to support both HDF4 and HDF5 as long as NASA needs them.
- Training and documentation, etc.
- Tools, etc. need multiple readers, writers

2. Interoperate HDF4 and HDF5

- Many environments will use older HDF4 and newer HDF5 together (e.g., data from Terra and Aqua)
  - This requires two reader/writers, one for HDF4 and one for HDF5.
  - Configuration is more complex
    - complex configure/make
    - potentially, very large binaries
    - For HDF-EOS: ***four***+ libraries

2. Interoperate HDF4 and HDF5

- Given software that uses HDFEOS4, it is usually not difficult to add HDFEOS5
- Many systems already support netCDF, GEOTiff, etc., supporting two versions of HDF-EOS is similar.
- The ‘heconvert’ utility uses both HDFEOS4 and HDFEOS5.

3. Convert data from HDF4 to HDF5

- In some cases, may want to convert data from HDF4 to HDF5 to work with new software:
  - on-demand, as needed
  - wholesale, e.g., whole collections

Our experiments show this is quite feasible.

(See below.)

3. Convert data from HDF4 to HDF5

- In many cases, will want a custom conversion, e.g., per data product
  - to capture specific semantics, e.g., relationships
  - to optimize the use of HDF5
  - to optimize the conversion, e.g., handle large objects, deal with missing values or compression, etc.

The NCSA libh4toh5 provides a toolkit to help build conversion utilities. (See below.)

II. Recent NCSA Activities

- NCSA has been establishing a toolkit to support transition
- Tools for Default Conversion of HDF4 to HDF5
  - Specification of default mappings
  - *h4toh5* utility – convert whole file to HDF5
  - *libh4toh5* – C library to convert individual objects

See: http://hdf.ncsa.uiuc.edu/h4-h5.html

Default Mapping

- Specification of default mapping of HDF4 and HDF5 concepts and objects
  - Conceptual guidance for developers
  - Standards to help interoperability
  - A general purpose solution, should be customized for particular uses

http://hdf.ncsa.uiuc.edu/HDF5/doc/ADGuide/H4toH5Mapping.pdf

Data Conversion Software

- *h4toh5* utility – conversion of 1 HDF4 file to HDF5
  - Implements the mapping specification
  - Model for customized conversion utilities
- *Libh4toh5* – Library of conversions for single objects, groups of objects
  - Same mapping and algorithms as *h4toh5* utility
  - Helps construction of custom conversions

For more info ant to obtain: http://hdf.ncsa.uiuc.edu/h4toh5

Feedback on these is requested!

Two Experiments

- Two Experiments testing conversion of HDF4 to HDF5 data
  - Used sample NASA datasets
- Experiment 1: Covert whole files with *h4toh5* utility
  - Default conversion of standard HDF
  - Didn't convert HDF-EOS objects
- Experiment 2: convert hybrid HDF-EOS files with *heconvert* plus *libh4toh5*
  - Convert HDF-EOS objects and other HDF objects

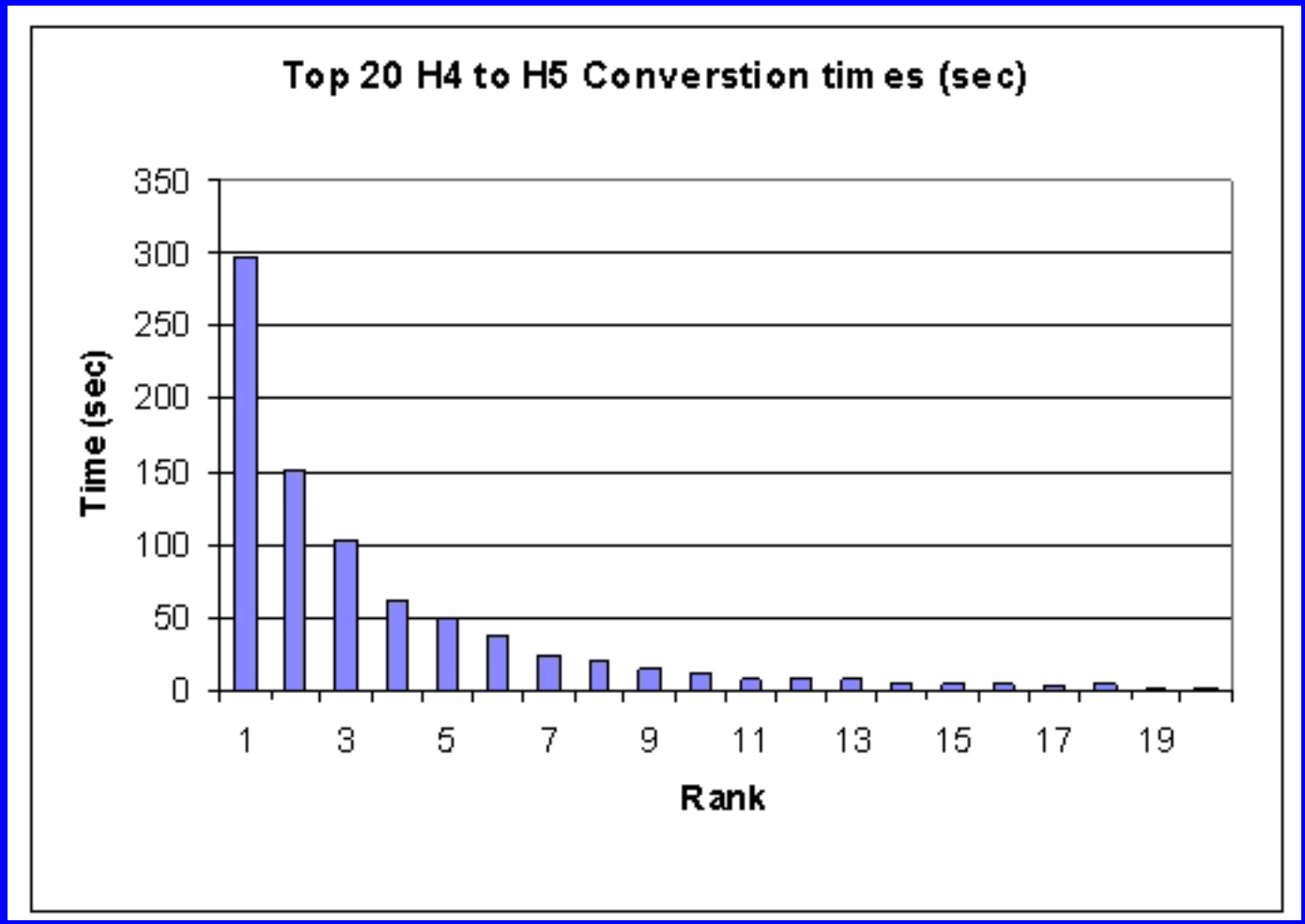
Conclusions

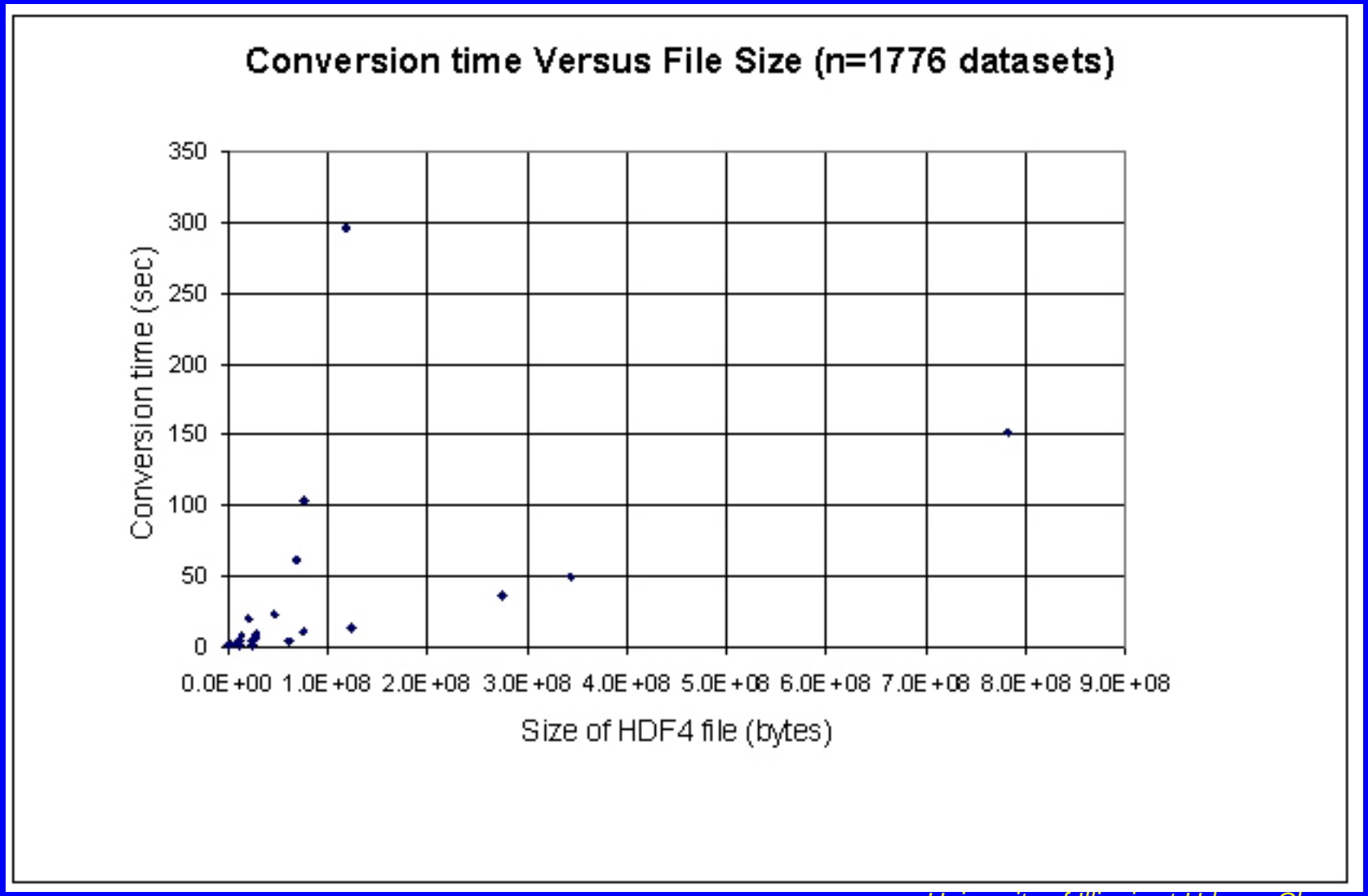
- Conversion from HDF4 to HDF5 is feasible, even for whole collections
  - Could be done on demand, or whole archives could be converted
- Custom product-specific conversion utilities are feasible
  - This is not particularly difficult
  - Demonstrates the effective use of HDFEOS4 and HDFEOS5 together

Experiment 1: conversion of NASA datasets to HDF5

- Tested with HDF4 files from NASA
  - Files from DIAL, DAACs
- All files converted successfully
- Conversion was fast

http://hdf.ncsa.uiuc.edu/h4toh5/Experiment/h4toh5exp.html


Top 20 H4 to H5 Converstion tim es (sec)
Time (sec)
350
300
250
200
150
100
50
0
1
3
5
7
9
11
13
15
17
19
Rank


Conversion time Versus File Size (n=1776 datasets)
Conversion time (sec)
350
300
250
200
150
100
50
0
0.0E+00
1.0E+08
2.0E+08
3.0E+08
4.0E+08
5.0E+08
6.0E+08
7.0E+08
8.0E+08
9.0E+08
Size of HDF4 file (bytes)

Conclusions

- For many NASA collections, conversion is feasible and fast
  - Could be done on demand, or whole archives could be converted

Caveats

- Not a representative sample of NASA data
- This is a *default* conversion: e.g., the result is not a valid HDF-EOS5 file
  - custom conversion would be desirable
- Files with very large individual objects performed much worse
  - newer products have this characteristic
  - conversion utility needs better memory management

Experiment 2: 'Hybrid' HDF-EOS files

- *heconvert* – utility that converts HDF-EOS4 to HDF-EOS5
  - Converts the HDF-EOS objects:
    - Grid, Swath, Point
    - StructMetadata.0
- Real EOS data are hybrids: they have other HDF4 objects
  - Vdata tables, SDS arrays
  - ProductMetadata.0, etc.
- These objects are omitted by *heconvert*

‘Hybrid’ HDF-EOS files

- ***Discussed this last year:*** almost all HDF-EOS files have standard HDF objects as well as HDF-EOS object. ***See previous talk (Workshop IV):***

  http://hdf.ncsa.uiuc.edu/h4toh5/ESDIS-sep/mcgrath-transition-issues/index.htm

- **Key issues**
  - Can convert the EOS objects
  - libh4toh5 can convert other objects individually
  - But is not trivial to find the ‘other’ objects
  - Probably requires custom conversion per data product

Experiment: 'Hybrid' HDF-EOS files

- Method: added calls to libh4toh5
  - Same conversion of EOS objects
  - Finds and converts (at least some) of the non-EOS objects
    - all Vdatas, annotations, images, 'lone' SDSs
- Data: files from the HDF-EOS Sampler CD
  - These files all have non-EOS objects

Results and Observations

- Results: libh4toh5 is a useful toolkit for constructing conversion tools
  - In most cases, all the objects are copied to the HDF5 file
  - Negligible overhead
- This is a ‘default’ conversion
  - Probably not the desired layout of the HDF5
  - Would want to use non-default options to libh4toh5, or possibly write custom

http://hdf.ncsa.uiuc.edu/h4toh5/Experiment2/heconvert.html

Summary

- HDFEOS4 and HDFEOS5 can be used in the same program
  - Can use existing and future data together
- Conversion of HDF4 to HDF5 is feasible
  - Conversion could be ‘on demand’ or for whole collections
- When conversion is desired, will likely want product-specific conversion
  - This is not particularly difficult
  - NCSA’s *libh4toh5* provides a good initial toolkit

URLs

- NCSA information page:
  http://hdf.ncsa.uiuc.edu/h4-h5/
- Conversion Utility Experiment
  http://hdf.ncsa.uiuc.edu/h4toh5/Experiment/h4toh5exp.html
- HE Conversion Experiment
  http://hdf.ncsa.uiuc.edu/h4toh5/Experiment2/heconvert.html